Opracowanie:
2006I
Stopnie powiązań

HISTORIA:

dokument systemu SINOL 1.6

1 Wprowadzenie

Postawione zadanie to klasyczny problem znajdowania najbardziej oddalonej pary wierzchołków w grafie (jego średnicy). Dla prostych grafów, o szczególnej strukturze, takich jak drzewa, można rozwiązać ten problem efektywniej, niż licząc odległości między wszystkimi parami wierzchołków — możemy obliczyć średnicę drzewa z użyciem programowania dynamicznego.

2 Rozwiązanie wzorcowe

Rozwiązanie wzorcowe szuka odległości między każdą parą wierzchołków za pomocą algorytmu Floyda–Warshalla. Złożoność czasowa tego algorytmu wynosi $O(n^3)$, natomiast pamięciowa $O(n^2)$. Implementacja nie wymaga zbyt wielu linii kodu co jest niewątpliwą zaletą.

3 Inne rozwiązania

3.1 Rozwiązania optymalne

Innym sposobem na rozwiązanie problemu w pesymistycznym czasie $O(n^3)$ jest wykonanie dla każdego wierzchołka przeszukiwania wszerz. Dzięki temu w każdym kroku znaleziona zostanie odległość pomiędzy wierzchołkiem początkowym a wszystkimi innymi wierzchołkami. Złożoność pamięciowa tego algorytmu wynosi $O(n)$, a więc ma przewagę nad rozwiązaniem wzorcowym. Sporą wadą takiego podejścia jest znacznie dłuższa i bardziej skomplikowana implementacja przez co łatwiej popełnić błąd.

3.2 Rozwiązania nieoptymalne

Ponieważ rozwiązanie optymalne jest dość intuicyjne, trudno mówić tu o rozwiązaniach nieoptymalnych, nie próbując jednocześnie w jakiś sztuczny sposób zwiększyć złożoności rozwiązania. Można zatem myśleć o rozwiązaniach, które będą obliczały odległość między każdą parą wierzchołków niezależnie, korzystając z algorytmu przeszukiwania wszerz (wówczas złożoność takiego rozwiązania wyniosłaby $O(n^4)$), czy nawet algorytmu Dijkstry (gdzie pesymistyczna złożoność mogłaby wynieść nawet $O(n^4 \cdot \log n)$).

3.3 Rozwiązania błędne

Ponieważ rozwiązanie oczywiste jest tym samym rozwiązaniem optymalnym, trudno jest mówić o rozwiązaniach błędnych. Wyobrazić sobie można natomiast scenariusz, w którym program znajduje odległość każdego wierzchołka do jednego danego (za pomocą algorytmu przeszukiwania wszerz), a następnie tylko na tej podstawie wnioskuje o maksymalnej odległości między parą wierzchołków w grafie.

4 Uwagi

Pierwszym krokiem w rozwiązaniu tego zadania jest zamiana reprezentacji grafu, zadanej poprzez imiona, na reprezentację liczbową, na której można w łatwiejszy sposób wykonywać operacje. Zatem przy wczytywaniu danych będziemy tworzyli on–line słownik odwzorowujący imiona osób, których sieć powiązań badamy, w liczby naturalne. Jeśli dane imię znajdziemy już w naszym słowniku, zwrócimy wcześniej przypisaną mu wartość, natomiast gdy będzie ona w tym słowniku nieobecna, dodamy ją do słownika, przydzielając jej kolejną liczbę naturalną. Z imionami i ich reprezentacją wiąże się jeszcze inny problem, gdyż w treści zadania nie ma podanych ograniczeń na długość pojedynczego imienia, czy też sumaryczną długość imion. Istotne jest więc aby wczytywać dane w sposób niezależny od długości imion, które potencjalnie mogą być bardzo duże.

5 Testy

Zdecydowaliśmy się stworzyć zestaw 10 testów. Każdy test składa się z grafów o określonej strukturze, z których żaden nie ma więcej niż $5 * numer\_testu$ wierzchołków.
W testach wyróżniamy podgrafy o następującej strukturze:

- ścieżki — ciąg wierzchołków połączony w jedną, długą ścieżkę, w ten sposób jesteśmy w stanie uzyskać maksymalne stopnie powiązań;
- cykle — tu podobnie jak w ścieżkach, jednak między każdymi dwoma wierzchołkami da się przejść na dwa sposoby, przez co maksymalny stopień powiązania nie przekracza połowy długości cyklu;
- drzewa — grafy nie posiadające cykli, które realizują średnie wartości stopni powiązań;
- kliki — grafy, w których każdy wierzchołek jest połączony z każdym innym przez co maksymalny stopień powiązań jest równy 1;
- grafy niespójne — inaczej nazywane w treści zadania rozdzielonymi, dla których maksymalny stopień powiązań jest równy ∞.

Często uzupełnione są one o losowe krawędzie, tworząc graf o zadanej gęstości. Wśród testów znajdują się także grafy zupełnie losowe, o zadanej gęstości, jednak wtedy z dużym prawdopodobieństwem graf będzie albo niespójny (jeśli gęstość będzie stosunkowo mała) albo będzie miał bardzo mały maksymalny stopnień powiązania (jeśli gęstość będzie większa).